



**HAL**  
open science

## Ecological forecasting models: Accuracy versus decisional quality

J.S. Pierre

► **To cite this version:**

J.S. Pierre. Ecological forecasting models: Accuracy versus decisional quality. *Ecological Modelling*, 2023, 482, pp.110392. 10.1016/j.ecolmodel.2023.110392 . hal-04123655

**HAL Id: hal-04123655**

**<https://hal.science/hal-04123655>**

Submitted on 21 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Ecological forecasting models: Accuracy *versus* decisional Quality

Jean-Sébastien Pierre(\*)

---

\* UMR #6553 Ecobio (Ecosystems, Biodiversity, Evolution) Université de Rennes 1 - CNRS, Campus scientifique de Beaulieu, 65042 Rennes cedex (France)

## 1. abstract

We consider here forecasting models in ecology or in agronomy, aiming at decision making based upon exceeding a quantitative threshold. We address specifically how to link the intrinsic quality of the model (its accuracy) with its decisional quality, ie its capacity to avoid false decisions and their associated costs. The accuracy of the model can be evaluated by the  $\rho$  of the regression of observed values versus estimated ones or by the determination coefficient  $R^2$ . We show that the decisional quality depends not only of this accuracy but also of the threshold retained to make the decision as well as on the state of nature. The two kinds of decisional errors consists either in deciding no action while an action is required (false negatives) or to act while it is useless (false positives). We also prove that the costs associated to those decisions depend also both of the accuracy of the model and of the value of the decision threshold.

## 2. Introduction

*framework*

Ecology, environment, crop protection, halieutics and wildlife management use more and more forecasting models of diverse nature: mechanistic deterministic models based upon ecosystemic processes knowledge, stochastic models analysing random functions, mixture of both such as SDE (Stochastic Differential equations) or statistic regression models obtained either by classical multiple regressions or by more sophisticated modern methods based on AI (Artificial Intelligence). Whatever the model's style, we address here the particular and widespread case, where the use of a forecasting model results in a decision of action based on a quantitative threshold: When the prediction of the model exceeds the threshold an action is decided, while in the inverse case, the decision is not to do anything. This procedure leads to

two types of decisional errors. The first one, analogous to false positives in clinical research (Scannell and Bosley, 2016<sup>[1]</sup>), is to decide an action while none is actually required. The reverse one, analogous to false negatives, is to decide of no action while one would be required. Those two types of errors are also well known in the statistical theory of tests, as the risks  $\alpha$  and  $\beta$  (Agosta, 2020<sup>[2]</sup>, Berger, 1985<sup>[3]</sup>, see also Figure 1). We show here that the probabilities of these two failures and their associated costs depend of the statistical quality of the model also called accuracy (Petchey *et al*, 2015<sup>[4]</sup>), of course, but also of the threshold accepted for the decision relatively to the states of nature. Actually, in the totally improbable case of a perfect model ( $R^2 = 1$ ) the probabilities of these two errors are null. In the general case of an imperfect model, however, it is obvious that the rarer the cases of threshold overshooting, the smaller is the probability of a positive decision as well as those of decisional errors and the associated costs. The present short note aims at quantifying this question. We restrict ourselves to the frequentist point of view, although the bayesian one could be of strong interest, especially if there is a need to question the model after a decision (Berger 1985, <sup>[3]</sup>, Williams and Hooten, 2016<sup>[5]</sup>).

#### *quality of ecological models*

We have to discuss in depth the concepts of statistical quality and of decisional quality and how they are linked. Most often, the statistical quality of the models are measured by standard statistics such as the  $R^2$ , the corrected  $R^2$  and various informational criteria derived from the Akaike informational Criterion (AIC, AICc, BIC, Mallows  $C_p$ , etc.). Planque *et al* <sup>[6]</sup> went a bit further, proposing a standard protocol to evaluate ecological models. Their analysis is very relevant but the question of the threshold is reduced to the question of defining the very threshold level (in the match between observed and modelled patterns) that can separate acceptable from unacceptable models. From another point of view, Grimm *et al*<sup>[7]</sup> present three questions to ask before using model output for decision support: what is the model's purpose? How is the model organized? Is there evidence the model works? We address specifically the latter that these authors precise by: What patterns can the model reproduce? Can the model make independent predictions, and under what conditions? We add also, at what cost? In what follows, we suppose that the two first questions are firmly set.

A general framework linking statistical inference to the decision theory was however set comprehensively by Williams and Hooten (2016)<sup>[5]</sup>. The case of forecasting models deserves a special consideration. Petchey *et al*(2015)<sup>[4]</sup> introduce the concept of ecological forecast horizon as the dimensional distance (in space or time) for which useful forecasts can be made. They also distinguish several important other concepts: accuracy, precision, uncertainty and forecast proficiency. Accuracy is the difference between an

observed and a predicted value, the actual criterion that we consider here. High accuracy implies good prediction and low accuracy poor prediction. Accuracy is an important component of forecast proficiency. We do not consider here neither their concept of precision nor that, closely linked of uncertainty because both are associated to the lack of knowledge about the processes modelled in the study. Actually, we take the forecasting model “as it is” judging only its performances on the decision process. Their definition of forecast proficiency is more fuzzy but leads to another useful concept : the forecast proficiency threshold, defined as the value of forecast proficiency above which forecasts are useful and below which forecasts are not useful.

All these concepts are of uttermost importance but do not address clearly the relationship between accuracy, decision errors and cost of the errors, what we try to do here. Chicco *et al* (2021)<sup>[8]</sup> adds another question: that of the definition of the cutoff threshold by maximizing the quality of the confusion matrix itself. Here we suppose that the cut off threshold is defined by external expert considerations, notably of cost. We do not discuss the merits of this threshold which is supposed to be set outside the modeling process.

#### *The confusion matrix*

Central for the decision theory is the object named confusion matrix (Table 1). On this basis, Fielding and Bell (1997)<sup>[9]</sup> discuss the nature of prediction errors and, subsequently, evaluates a range of techniques that may be used to assess and compare prediction success. Their work focuses on the particular case of presence absence of a species in an area and defines – as in clinical research – two types of errors : the False Positive (FP) – the species is decided present while it is not -, and the False Negative (FN) – the species is decided absent while it exists in the area of interest. They note that “The ecological literature seems to have paid little attention to how the partitioning method can influence the error rates”. So they address the question of the partitioning model accuracy. In the figure 2 of their work, they address a problem very similar to our, the influence of the cut-point threshold of the three error rates (False Positive Rate, False Negative Rate and Total misclassification rate. They also evoke the concept of cost matrices.

Given those definitions and limitations, we analyse the consequences of the accuracy of the model on the quality of decision, along with the associated costs. This will be done on the bases of the confusion matrix as the major criterion of the decision quality.

### **3. The problem**

Let us suppose that we have at hand a predictive model which was validated by observations realised in the true life. Its quality can be evaluated

by a statistical relation between predicted and observed data. We suppose in what follows that the relation is linear, unbiased (slope = 1) and without intercept (no intercept bias). Let us note  $X$  the bivector  $(x, y)$  where  $x$  is the set of predicted values and  $y$  the set of observed values. One expects often that the vector  $X$  follows a bivariate normal distribution whose probability density function is:

$$\Phi(X) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}[\dot{X}^T \Sigma^{-1} \dot{X}]} \quad (1)$$

where  $\dot{X}$  is the z-score transform of the vector  $X = \begin{pmatrix} x \\ y \end{pmatrix}$ ,  $\dot{X}^T$  is its transpose (row vector),  $|\Sigma|$ , the determinant of  $\Sigma$ , equals  $\sigma_x \sigma_y \sqrt{1 - \rho^2}$ . Of course  $x$  is calculated on each case before the realisation of the phenomenon, and  $y$  is observed later for each value of  $x$ .

#### *Some possible generalisations*

Other bivariate distributions may be considered in place of the normal one. Most of continuous variables may be approximatively normalised by transformations (logarithmic, square root, arc sine of square root, etc.). Others can be used directly (beta, weibull). If predicted and observed values are counts, a bivariate poisson distribution with or without overdispersion (Karlis and Ntzoufras 2005) <sup>[10]</sup> can be used. In the case where the variable is a ratio of counts the bivariate binomial distribution (Biswas and Hwang 2002) <sup>[11]</sup> is adequate. We will not detail all the possible cases here but many bivariate distributions are now available in the statistical literature.

#### *Decision*

The threshold  $A$  define two states of nature: the variable to predict will be either really lower to  $A$  or really greater or equal to  $A$  (Table 1). One has to decide an action  $\mathcal{A}$  when the predicted value is greater or equal to the  $A$ . The decision scheme is:

$$\begin{aligned} x \geq A &\Rightarrow \mathcal{A} \\ x < A &\Rightarrow |\mathcal{A} \end{aligned} \quad (2)$$

$|\mathcal{A}$  standing for "not  $\mathcal{A}$ ". The correct decisions are taken in the two following cases:

$$\begin{aligned} x \geq A \wedge y \geq A &\Rightarrow \mathcal{A} \\ x < A \wedge y < A &\Rightarrow |\mathcal{A} \end{aligned} \quad (3)$$

And two erroneous decisions can occur when:

$$\begin{aligned} x \geq A \wedge y < A &\Rightarrow \mathcal{A} \\ x < A \wedge y \geq A &\Rightarrow |\mathcal{A} \end{aligned} \quad (4)$$

The first equation case (refeq4) is the equivalent of a false positive in clinical research. This results in the application of the *mathcal{A}* action while

*mathcal{A}* is useless. The second equation case (refeq4) is the equivalent of a false negative, it results in not applying *mathcal{A}* while *mathcal{A}* was needed. Except when high costs are involved in applying the *mathcal{A}* action, (see §3) the second case is generally considered more damaging than the first, a reason why we will call the case 1 "critical error". The set of decisions taken with reference to a model organises as a special case of the so-called "confusion matrix" well known in the decision theory (Agosta 2020) reffibagosta (Table reftab1).

		state of nature	
		needed	not needed
decision	intervention	correct decision	<b>false positive</b>
	no intervention	<b>false negative (Critical error)</b>	correct decision

Table 1: Confusion matrix

Figure 1 illustrates the four possible outcomes. As the overall integral of the bivariate normal distribution sums to 1, the integrals of  $\Phi(.,.)$  respectively in the red and blue hatched areas give directly the probabilities of both type of errors, and the sum of the two gives the total probability of a decision error. More precisely:

$$\begin{aligned}
 P_1 &= P(\mathcal{A} | \neg \mathcal{A}) = \int_A^\infty \int_{-\infty}^A \Phi(x, y) dy dx \\
 P_2 &= P(\neg \mathcal{A} | \mathcal{A}) = \int_{-\infty}^A \int_A^\infty \Phi(x, y) dy dx \\
 P_{err} &= P_1 + P_2
 \end{aligned} \tag{5}$$

Let us note that, by symmetry, we should have  $P_1 = P_2$  and thus  $P_{err} = 2P_1 = 2P_2$ . As the cumulative probability functions of the normal distribution have no close form expression, the two double integrals corresponding to red and blue hatched areas of figure 1 require to be calculated numerically. Quite a lot of numerical integrators are currently available on line, here we used the function `ndpdf` of the package `binomial` of *R* (*R* core team 2022) [12]. These expressions are therefore suitable to plot the errors rates which occur for a set of thresholds  $A$  and a set of correlation coefficients  $\rho$  or of determination coefficients  $R^2$  representing better the overall statistical quality of the model. This is shown in Figure 2.

In the case of counts and of bivariate discrete distribution, considering  $\Phi(.,.)$  not as a probability density but as a discrete probability function  $\Phi(m, n)$ , equation (5) holds by replacing integrals by discrete sums and the left border  $-\infty$  by 0.

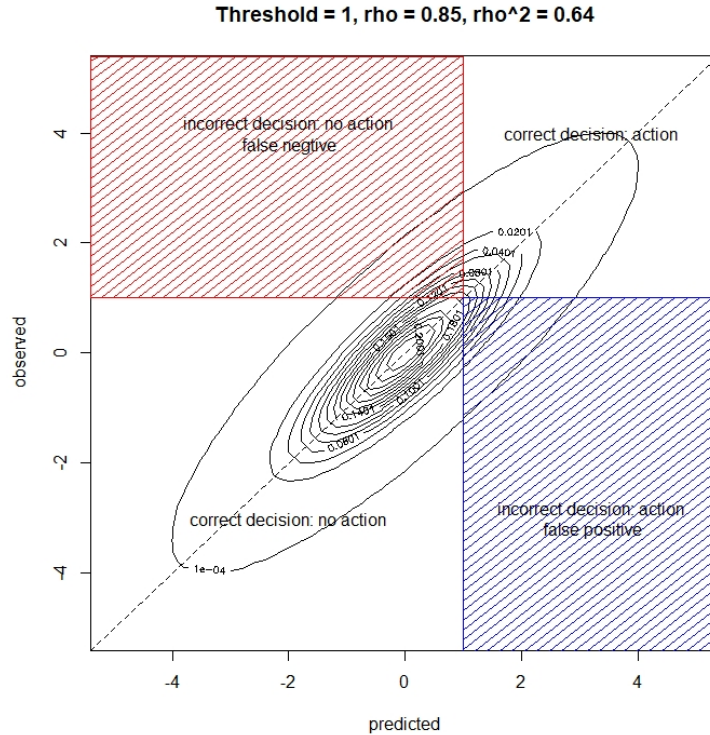


Figure 1: The outcomes of the taken decisions with an unbiased model; hatched in red: false negative error, hatched in blue: false positive error. Not hatched: correct decisions. The bivariate normal distribution is figured by isodensity lines. Estimated and observed values are z-score transformed and thus zero in horizontal and vertical axis represent the means and the tickmarks are graduated in standard deviation units.

#### 4. Rates of Error as a function of the threshold of decision and model quality

Equations (5) allow to calculate the rate of decision errors obtained for a given quality  $\rho$  of the model and a given threshold of decision  $A$ .

We restrict our analysis to the case where  $\rho > 0$  because it is the normal situation for the regression of observed versus estimated values in a model. We also restrict our analysis to thresholds of decision greater than the mean, because it is unlikely to work with very low threshold of decisions. Figure 2 shows four representations of the surface obtained by varying both  $\rho$  and  $A$ .

The figure shows that the rate of decision error is very sensitive to the chosen threshold. When the threshold is near the mean, a very high statistical quality of the model is required to get an acceptable rate of decision error. At the contrary, for high thresholds, greater than 2.5 standard deviations, the probability of decision error is very low, regardless of the model quality which has few influence. Of course, when the model is little informative, *i.e.*

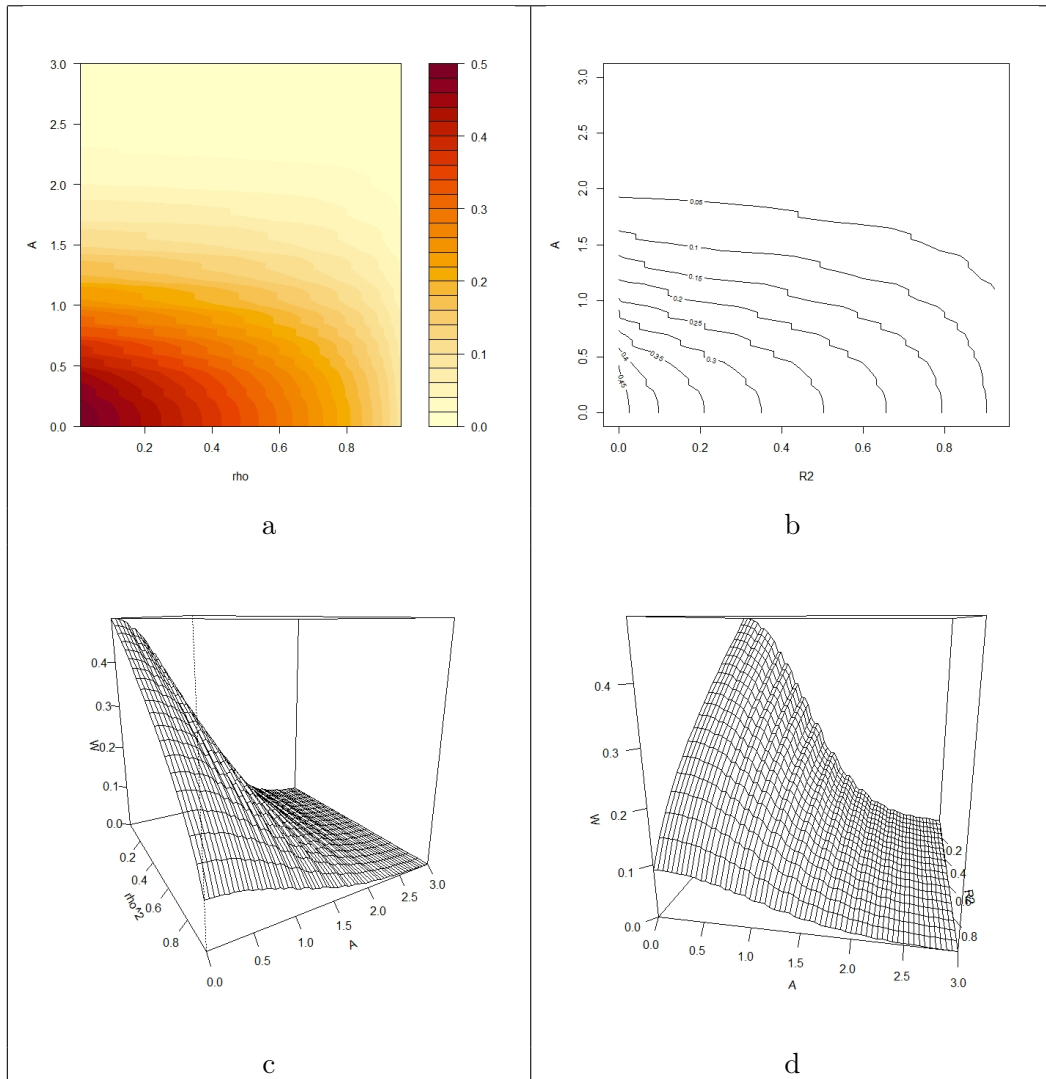


Figure 2: Error rates occurring for a model of quality  $\rho$  and a threshold of decision  $A$ ; a: coloured contour plot, b: black and white contour plot with isolines values, c: perspective plot: profile view, d: perspective plot: front view.

$\rho \simeq 0$  and when the threshold is close to the mean of predicted values, the rate of decision error is close to 50%.

## 5. Consequences on the risk and losses

Here, we follow the definition of loss and risk adopted by Williams and Hooten (2016)<sup>[5]</sup>. Suppose that one decision has to be made, let us define:

- $c_1$  the cost of an intervention (or harmfulness threshold)



- $c_2$  the cost of the damages in case of false negative

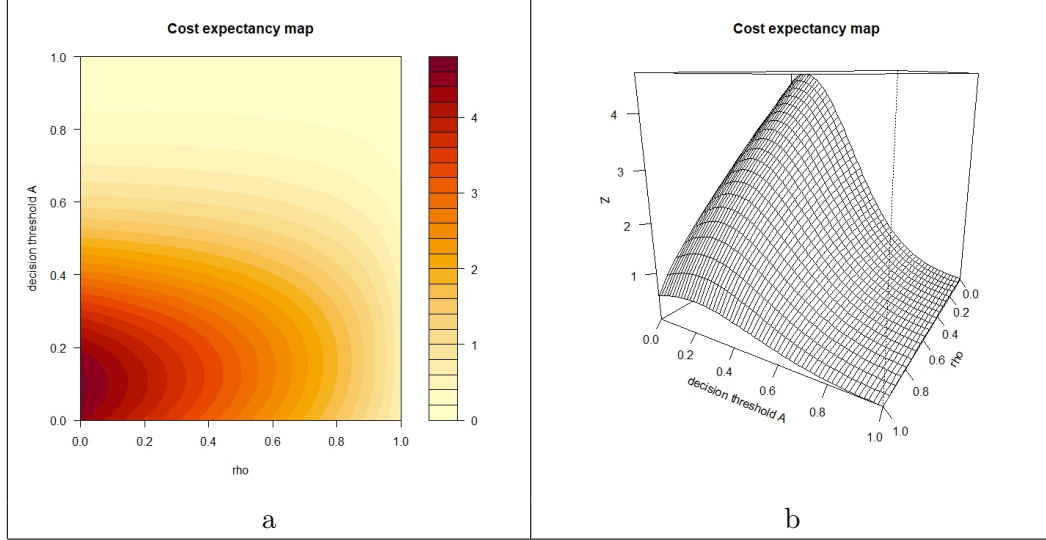


Figure 3: Map of error costs: the decisions are either to intervene or not to intervene according to the threshold  $A$ . Costs for a model of quality  $\rho$  and a threshold of decision  $A$ ; a: coloured contour plot, b: perspective plot.

Furthermore, we suppose that  $c_2$  is an increasing function of the realised phenomenon  $y$ . For sake of simplicity let us set  $c_2 = c_1 + \alpha y$ .  $c_1$  is natural as a basis because it is generally admitted that the threshold should be set at the point where the cost of the damages suffered equals the cost of the intervention (called the harmfulness threshold in the case of plants pests and plant diseases and other fields of research. (See *inter alia* Bossis, 2020,<sup>[13]</sup>, Colbach and Cordeau, 2018,<sup>[14]</sup>, Rusu, 2010<sup>[15]</sup>, Maslennikov *et al.*, 2018<sup>[16]</sup>, Wiekopolan *et al.*, 2018<sup>[17]</sup>).

We have to add to equation (5) the probability of taking the correct decision of an intervention:

$$P_3 = P(\mathcal{A}|\mathcal{A}) = \int_A^\infty \int_A^\infty \Phi(x, y) dy dx \quad (6)$$

In terms of loss, the risk becomes dissymmetric, and can be set as:

$$E(C) = P_1 c_1 + \int_{-\infty}^A \int_A^\infty \Phi(x, y) (c_1 + \alpha y) dy dx \quad (7)$$

$P_1$  and  $P_3$  as defined by equations (5) and (6). The last integral is the convolution of  $P_2$  and of the loss induced by each possible  $y$  actually greater than  $A$ .

This is the cost and needs also to be integrated numerically (Figure3). This holds for the bivariate normal distribution. In the case of counts and

of bivariate discrete distribution, equations (6) and(7) hold by replacing integrals by discrete sums and the left border  $-\infty$  by 0. Figure 3 shows some results obtained with  $c_1 = 10$  (conventional units) and  $\alpha = 10$ . These maps are *a priori* maps for the model and shows the cost (expected loss) of the decisions it implies. For low thresholds  $A$ , the risk of decisions decrease very slowly with the quality of the models, showing that the lower the threshold, the better should be the statistical quality of the model.

## 6. Conclusion

We present here a framework to judge the decisional quality of forecasting models from their statistical quality, also called accuracy. It can be seen as an adaptation to ecology and agronomy of the quality tests used in clinical research (Scannel and Bosley 2016)<sup>[1]</sup>, while adding cost considerations which are of outstanding importance in environmental researches. The quality of decision making depends strongly of two parameters, and not of a single one. Excepted for a very high statistical quality of the model, a decision threshold close to the mean of the phenomenon under study induces a large rate of decision error, and a high (frequentist) risk rate (see Figure 2a). The problem is that the decision threshold does not depend at all of the model, but only of the costs and benefits associated with its use. As an example, in crop protection procedures, the threshold required to decide an intervention (e.g. insecticide spraying or biocontrol) is generally defined as a defined population density of a pest or as a defined prevalence of the disease, such that the costs of the damages suffered by the crop overcome the cost of the intervention (harmfulness threshold). More and more the mere financial estimation of the damages and of the interventions must be combined with the ecological impact of them. Our calculations show that the decision taken will be all the more relevant as the threshold of decision is higher, compared to the average situation. In other words, when the damages are rare, the decision support system and the model building are easier and more relevant than in situations where damages are of high probability. The closer of the mean is the intervention threshold, the more accurate must be the model. All these considerations are of little help in the model building process, but could give precious indications on the interest to apply it, its "proficiency" (Petchey *et al*, 2015)<sup>[4]</sup>. Let us note that using the model or not is itself a decision problem, that we do not treat here but for which the realisation of the maps of Figure 3 may be a help.

## 7. Acknowledgements

We thank warmly Nicolas Parisey and Eric Wajnberg for their help in revising this article. We thank also the two anonymous referees who gave their useful comments on the first version.

## References

- [1] Scannell, J. W., and J. Bosley. 2016, When Quality Beats Quantity: Decision Theory, Drug Discovery, and the Reproducibility Crisis. PLoS ONE 11:e0147215.
- [2] Agosta J.M. 2020, <https://medium.com/decision-analysis/decision-quality-for-data-scientists-lessons-from-decision-professionals-5a52100a7b41>
- [3] Berger, J. O. 1985, Statistical decision theory and Bayesian analysis. Springer, New York, New York, USA.
- [4] Petchey O.L., Pontarp M.M., Thomas M., Kéfi S., Ozgul A., Weilenmann M., Palamara G. M., Altermatt F.M.B., and Levine, J.M. 2015, The ecological forecast horizon, and examples of its uses and determinants Ecology letters, 18,7,597-611
- [5] Williams P. J., and Hooten M.B. 2016, Combining statistical inference and decisions in ecology. Ecological Applications 26:1930-1942.
- [6] Planque B., Aarflot J.M., Buttay L., Carroll J., Fransner F., Hansen C., Husson B., Langangen Ø., Lindstrøm U., Pedersen T., Primicerio R., Sivel E., Skogen M.D., Strombom E., Stige L.C., Varpe Ø., Yoccoz N.G., 2022, A standard protocol for describing the evaluation of ecological models. Ecological Modelling 471, 110059.
- [7] Grimm V., Johnston A.S.A., Thulke H.H., Forbes V.E., Thorbek P., 2020. Three questions to ask before using model outputs for decision support. Nature Communications 11, 4959.
- [8] Chicco D., Totsch N., Jurman G., 2021, The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. Biodata Mining 14.
- [9] Fielding, A. H. and Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24(1): 38-49.
- [10] Karlis D. and Ntzoufras I. 2005, Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R Journal of Statistical Software Vol. 14 Issue 10 36p
- [11] Biswas, A. and J.-S. Hwang 2002, A new bivariate binomial distribution. Statistics & Probability Letters 60(2): 231-240.
- [12] R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

- [13] Bossis, M., 2020. Control of *Heterodera carotae* Jones (1950): Effectiveness of available methods and prospects, Progress on Pest Management in Field Vegetables. CRC Press, pp. 187-193.
- [14] Colbach, N., Cordeau, S., 2018. Reduced herbicide use does not increase crop yield loss if it is compensated by alternative preventive and curative measures. European Journal of Agronomy 94, 67-78.
- [15] Rusu, G., 2010. Harmfulness threshold produced by unsemelling camomile (matricaria inodora l.) in wheat winter crop. Scientific Papers-University of Agronomic Sciences and Veterinary Medicine Bucharest. Series A, Agronomy 53, 308-311.
- [16] Maslennikov, A., Demidova, S., Ryabova, A., 2018. Ecological and toxicological assessment of hazard, posed by water contamination of water bodies with polyvinyl nitrate. Toxicological Review, 12-19.
- [17] Wielkopolan, B., Baran, M., Roik, K., 2018. The insecticide application effect on combating of pea weevil larvae (*Bruchus pisorum* L.) during the cultivation of Milwa peas. Przemysl Chemiczny 97, 1316-1319.